NASA CONTRACTOR REPORT

NASA CR-

c. 1

NASA CR-983

QUANTUM MECHANICAL STUDY OF MOLECULES

Eigenvalues and Eigenvectors of Real Symmetric Matrices

by G. R. Verma, C. D. La Budde, and R. C. Sahni

Prepared by
NEW YORK UNIVERSITY
University Heights, N. Y.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • DECEMBER 1967

NASA CR-983

QUANTUM MECHANICAL STUDY OF MOLECULES

Eigenvalues and Eigenvectors of Real Symmetric Matrices

By G. R. Verma, C. D. La Budde, and R. C. Sahni

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Prepared under Grant No. NsG-76 by
NEW YORK UNIVERSITY
University Heights, N.Y.

for

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

QUANTUM MECHANICAL STUDY OF MOLECULES

Eigenvalues and Eigenvectors of Real Symmetric Matrices

by G. R. Verma,* C. D. La Budde and R. C. Sahni

SUMMARY

In this report, three general classes of methods for calculating the eigenvalues and eigenvectors of real symmetric matrices arising in quantum mechanical calculations are described: the Sturm sequence methods, the orthogonal reduction methods, and the step by step methods. The advantages and limitations of each method are pointed out. The report also includes the discussion of various methods of reducing real symmetric matrices to more compact convenient forms. Methods of reduction to tridiagonal form, and deflation of matrices to smaller order are described.

INTRODUCTION

This is the first of a series of reports on the present techniques used to perform matrix calculations on high speed electronic digital computers. In this report the authors confine themselves to a discussion of techniques for calculating the eigenvalue and eigenvectors of real symmetric matrices arising in problems of molecular quantum mechanics. Subsequent reports will deal with other numerical techniques used in solving problems arising in molecular quantum mechanics.

1. Statement of the Problem

For a given square nxn matrix $A = (\alpha_{ij})$, find numbers (eigenvalues) ϵ and non zero column vectors (eigenvectors) X such that

$$AX = \epsilon X \tag{1}$$

where A is real (all α_{ij} real) and symmetric ($\alpha_{ij} = \alpha_{ji}$ for all i, j). We know

* Presently at the University of Rhode Island, Kingston, Rhode Island

by the theory of linear equations there exist non-zero X's satisfying (1) only if the determinant of A' (written DET (A')) is zero where A' has elements $\alpha'_{ij} = -\alpha_{ij}$ $i \neq j$ and $\alpha'_{ii} = \epsilon - \alpha_{ii}$. DET (A'), as a function of ϵ is an n^{th} degree polynomial in ϵ, and is referred to as the characteristic polynomial of A. We denote DET (A') by $f(\epsilon)$.

If we let I represent the identity matrix, A^{-1}, A^{T} represent the inverse and transpose of a matrix A respectively, then we may state two eigenvalue problems closely related to the one given by equation (1) namely

$$AX = \epsilon BX \tag{2}$$

and

$$ABX = \epsilon X \tag{3}$$

where A and B are symmetric matrices and B is, in addition, positive definite (all eigenvalues of B are positive).

If B is symmetric and positive definite, then there is an invertible matrix L (L^{-1} exists) such that $L^{T}L = B$. By the theory of linear equations the ϵ's of equation (2) must satisfy

$$0 = \mathrm{DET}\ (A - \epsilon B) \tag{4}$$

or

$$\begin{aligned} 0 &= \mathrm{DET}\ ((L^{T})^{-1})\ \mathrm{DET}\ (A - \epsilon B)\ \mathrm{DET}\ (L^{-1}) \\ &= \mathrm{DET}\ ((L^{T})^{-1}\ (A - \epsilon B)\ L^{-1}) \qquad (5) \\ &= \mathrm{DET}\ ((L^{T})^{-1}\ AL^{-1} - \epsilon I) \\ &= \mathrm{DET}\ (C - \epsilon I) \end{aligned}$$

where $C = (L^{T})^{-1}\ AL^{-1}$ and is also symmetric. Hence the solution of (2) is equivalent to the solution of

$$CY = \epsilon Y \tag{6}$$

where the ϵ are the eigenvalues of C and the Y are the corresponding eigenvectors. The vectors Y are related to the vectors X of equation (2) as follows:

$$X = L^{-1}Y. \tag{7}$$

This can be verified by substitution in equation (2) and multiplication on the left by $(L^T)^{-1}$.

The ϵ of equation (3) must satisfy

$$\mathrm{DET}\ (AB - \epsilon I) = 0 \tag{8}$$

or

$$\begin{aligned} 0 &= \mathrm{DET}\ (AB - \epsilon I)\ \mathrm{DET}\ (B^{-1}) \qquad (9) \\ &= \mathrm{DET}\ (A - \epsilon B^{-1}). \end{aligned}$$

This is essentially equation (4) and the eigenvalue problem can be put in the form

$$CY = \epsilon Y \tag{10}$$

where $C = LAL^T$ and $X = LY$ in equation (3).

2. Description and Classification of the Methods

There are three general classes of numerical procedures for solving the basic problem stated in equation (1) of section 1.

The first class of methods, referred to as the Sturm sequence methods, determines the numbers ϵ by means of a Sturm sequence of polynomials associated with the matrix A. Once the ϵ have been determined, the associated vectors X can be determined in several straight forward methods which will be described later.

The second class of methods, referred to as the orthogonal reduction methods, determines the numbers ϵ and vectors X simultaneously. A sequence of orthogonal matrices $U^{(k)}$ is generated, usually as a product of elementary orthogonal matrices, such that the limit as $k \longrightarrow \infty$ of $U^{(k)T} AU^{(k)}$ is a diagonal matrix (zeros everywhere except possibly on the main diagonal). The numbers on the diagonal will be the eigenvalues ϵ of A and the corresponding columns of $U =$ limit of $U^{(k)}$ as $K \longrightarrow \infty$ will be the eigenvectors X.

The third class of methods, referred to as the step by step methods, determines one eigenvalue ϵ and a corresponding eigenvector X, one pair, ϵ and X at a time. Once ϵ and X have been determined, it is easy to construct an orthogonal matrix V such that

$$V^T AV = \left(\begin{array}{c|c} \epsilon & 0 \\ \hline 0 & A_1 \end{array} \right) \qquad (11)$$

where A_1 is a symmetric (n-1) X (n-1) matrix. Thus, the problem has been essentially reduced by reducing the order of the matrix by one.

2.1 Two Special Matrices

We will now define two special orthogonal matrices which will be used in a number of algorithms described in this report.

Let J_{ij} $(i < j)$ be an nxn matrix of the following form

$$J_{ij} = \begin{array}{c} \\ \\ \\ \text{row } i \\ \\ \\ \text{row } j \\ \\ \\ \\ \end{array} \begin{array}{c} \begin{array}{cccccccc} & & \text{col } i & & & \text{col } j & & \end{array} \\ \left(\begin{array}{cccccccc} 1 & & & & & & \bigcirc & \\ & 1 & & & & & & \\ & & c & & & s & & \\ & & & 1 & & & & \\ & & & & 1 & & & \\ & & s & & & -c & & \\ & & & & & & 1 & \\ & \bigcirc & & & & & & \ddots \\ & & & & & & & 1 \end{array} \right) \end{array} . \qquad (12)$$

J_{ij} is the same as the identity matrix except for the i^{th} row i^{th} column, i^{th} row j^{th} column, j^{th} row i^{th} column, j^{th} row j^{th} column where the entries are c, s, s, and -c respectively. If

$$c^2 + s^2 = 1 \tag{13}$$

then J_{ij} is both orthogonal and symmetric. We shall call the matrices J_{ij} Jacobi plane rotations or Jacobi transformations (matrices).

Let

$$H_i = I - 2W_iW_i^T \tag{14}$$

be a matrix with W_i^T a row vector having the property that the first i components of W_i^T are zeros, namely

$$W_i^T = (0, 0, \ldots, 0, w_{i+1}, \ldots, w_n). \tag{15}$$

If

$$W_i^TW_i = 1 \tag{16}$$

then H_i is both orthogonal and symmetric. We shall call the matrices H_i Householder Transformations Matrices. It is easily seen that H_{n-2} is equal to $J_{n-1,n}$ for some suitably chosen c, s.

2.2 Preliminary Reductions

The solution of the general problem of finding an orthogonal matrix U for a given symmetric matrix A such that $U^T AU = D$ where D is a diagonal matrix is often facilitated by making some preliminary reductions by orthogonal similarity transformations on A. We can, for instance, reduce a symmetric matrix to tri-diagonal (Jacobi) form (zeros everywhere except on the three main diagonals) by several non-iterative methods, three of which will be described here. If an eigenvalue and a corresponding eigenvector of the nxn matrix is known, we may use this information to transform the matrix into a direct sum of a n x n matrix and an (n-1) x (n-1) matrix which essentially reduces the order of the matrix to be solved. This process is referred to as deflation. We will describe here the reduction to tridiagonal form by the methods of Givens, Householder, and Lanczos, and one of the deflation methods.

We will first describe Givens' method. Suppose A is a symmetric matrix of the form

$$A = \left(\begin{array}{cc:c} \multicolumn{2}{c:}{A_i} & \begin{array}{c} 0 \\ \hdashline B_{n-i}^T \end{array} \\ \hdashline 0 & B_{n-i} & A_{n-i} \end{array}\right) \tag{17}$$

where A_i and A_{n-i} are i x i and (n-i) x (n-i) matrices respectively, A_i is in tridiagonal form, and B_{n-i} is an (n-i) dimensional column vector. We wish to show how to obtain an orthogonal matrix U such that $U^TAU = A'$ where

$$A' = \left(\begin{array}{cc:c} \multicolumn{2}{c:}{A_i'} & \begin{array}{c} 0 \\ \hdashline B_{n-i}'^T \end{array} \\ \hdashline 0 & B_{n-i}' & A_{n-i}' \end{array}\right) \tag{18}$$

and A_i', A_{n-i}' are i x i and (n-i) x (n-i) matrices where A_i' is in tridiagonal form and B_{n-i}' is an (n-i) dimensional vector having all components except possibly the first equal to zero. Then A' will be of the form

$$A' = \left(\begin{array}{c|c} A'_{i+1} & \begin{array}{c} 0 \\ \hline B'^{T}_{n-i-1} \end{array} \\ \hline \begin{array}{c|c} 0 & B'_{n-i-1} \end{array} & A'_{n-i-1} \end{array}\right) \tag{19}$$

where A'_{i+1}, A'_{n-i-1} are $(i+1) \times (i+1)$ and $(n-i-1) \times (n-i-1)$ matrices respectively, A'_{i+1} is in tridiagonal form and B'_{n-i-1} is an $(n-i-1)$ dimensional vector. We consider a sequence of $(n-i-1)$ Jacobi transformations $J^{(1)}_{n-1,n}$, $J^{(2)}_{n-2,n-1}$, - - - - - $J^{(k)}_{n-k-1,n-k}$, $J^{(n-i-1)}_{i+1,i+2}$. Let

$$\begin{aligned}
&A^{(0)} = A \\
&J^{(1)}_{n-1,n} \; A^{(0)} \; J^{(1)}_{n-1,n} = A^{(1)} \\
&J^{(k)}_{n-k-1,\, n-k} \; A^{(k-1)} \; J^{(k)}_{n-k-1,\, n-k} = A^{(k)} \\
&A^{(n-i-1)} \\
&B^{(0)T}_{n-i} = B^{T}_{n-i} = (\alpha_1, \alpha_2, \ldots, \alpha_{n-i}) \\
&B^{(k)T}_{n-i} = (\alpha^{(k)}_1, \alpha^{(k)}_2, \ldots, \alpha^{(k)}_{n-i}) \\
&B^{(n-i-1)T}_{n-i} = B'^{T}_{n-i} = (\alpha'_1, \alpha'_2, \ldots, \alpha'_{n-i}) \; .
\end{aligned} \tag{20}$$

Then we define

$$U^{T} = J^{(n-i-1)}_{i+1,\, i+2} \; ---- \; J^{(2)}_{n-2,\, n-1} \; J^{(1)}_{n-1,n} \; . \tag{21}$$

The transformations $J^{(k)}_{n-k-1,\, n-k}$ will not destroy the tridiagonal form of A_i, namely $A'_i = A_i$. The blocks of zeros in the upper right and lower left hand corners of A will be preserved by these transformations. We need only consider the effects of these transformations on the vectors $B^{(k)}_{n-i}$. Let the c, s associated with the $J^{(k)}_{n-k-1,\, n-k}$ be denoted by $c^{(k)}$, $s^{(k)}$. If we define

$$s^{(k)} = \alpha^{(k-1)}_{n-i-k+1} \Big/ \left[(\alpha^{(k-1)}_{n-i-k+1})^2 + (\alpha^{(k-1)}_{n-i-k})^2\right]^{1/2} \qquad (22)$$

$$c^{(k)} = \alpha^{(k-1)}_{n-i-k} \Big/ \left[(\alpha^{(k-1)}_{n-i-k+1})^2 + (\alpha^{(k-1)}_{n-i-k})^2\right]^{1/2}$$

then $(s^{(k)})^2 + (c^{(k)})^2 = 1$ and $\alpha^{(k)}_{n-i-k+1} = 0$.

If $\alpha^{(k-1)}_{n-i-k+2} = \alpha^{(k-1)}_{n-i-k+3} = \cdots\cdots = \alpha^{(k-1)}_{n-i} = 0$, then

$\alpha^{(k)}_{n-i-k+2} = \alpha^{(k)}_{n-i-k+3} = \cdots\cdots = \alpha^{(k)}_{n-i} = 0$ also. Thus it can be easily seen that B'_{n-1} will have all components equal to zero, except possibly the first.

The Householder reduction to tridiagonal form is as follows. Suppose A has the form of equation (17) and let

$$A' = H_i \, AH_i. \qquad (23)$$

We wish to show that for a suitably chosen W_i, A' will be in the form of equation (19). Let $W_i^T = (o, \ldots, o, w_{i+1}, \ldots, w_n)$ be defined as follows

$$w_{i+1} = \left[1/2 \; (1 + |\alpha_1| \; (\sum_{k=1}^{n-i} \alpha_k^2)^{-1/2} \right]^{1/2}$$

$$(24)$$

$$w_\ell = \mathrm{sgn}(\alpha_1) \, \alpha_{\ell-i} \Big/ \left[(\sum_{k=1}^{n-i} \alpha_k^2)^{1/2} \, w_{i+1} \right], \; i+2 < \ell < n$$

It is a matter of straight forward algebra to verify that $W_i^T \, W_i = 1$ and that the vector B'_{n-i} has all components equal to zero except possibly the first. It is also easily seen that $A_i' = A_i$ and that the upper right and lower left hand blocks of zeros in equation (17) are preserved.

The Lanczos method may be described as follows. Let X_1 be any arbitrary non-zero vector. We define the following sequence of vectors X_i

$$\begin{aligned} X_2 &= AX_1 - \alpha_1 X_1 \\ X_3 &= AX_2 - \alpha_2 X_2 - \beta_1 X_1 \\ X_i &= AX_{i-1} - \alpha_{i-1} X_{i-1} - \beta_{i-2} X_{i-2} \end{aligned} \tag{25}$$

where

$$\begin{aligned} \alpha_{i-1} &= (X_{i-1}^T AX_{i-1})/(X_{i-1}^T X_{i-1}) \\ \beta_{i-2} &= (X_{i-1}^T AX_{i-2})/(X_{i-2}^T X_{i-2}) . \end{aligned} \tag{26}$$

It can easily be seen that the vectors X_i are orthogonal to each other. We distinguish two cases: (1) $X_j^T X_j \neq 0 \;\; 1 \leq j \leq i-1$ and $X_i^T X_i = 0$ (or $X_i = 0$) for some i, $2 \leq i \leq n$ or (2) $X_j^T X_j \neq 0$ for $1 \leq j \leq n$.

Case (1) $X_j^T X_j \neq 0 \;\; 1 \leq j \leq i - 1$; $X_i = 0$ for some i. Then we have

$$AX_{i-1} = \alpha_{i-1} X_{i-1} + \beta_{i-2} X_{i-2} \tag{27}$$

and the space spanned by the orthogonal vectors $X_1, \ldots, X_{i-1}$ is a reducing subspace for A. We define an nxn matrix V

$$V = (X_1, X_2, X_3, \ldots, X_{i-1}, Y_i, \ldots, Y_n) \tag{28}$$

where the Y_j s are column vectors so chosen that $X_j^T Y_k = 0$ and $Y_j^T Y_k = \delta_{jk}$.

$$V^T V = D \tag{29}$$

where D is a diagonal matrix with positive elements on the diagonal. Hence D has a real square root $D^{1/2}$ and $D^{-1/2}$ exists. Then there is a matrix A'

$$A' = \left(\begin{array}{c|c} Z_{i-1} & 0 \\ \hline 0 & Z_{n-i+1} \end{array}\right) \tag{30}$$

where Z_{i-1} and Z_{n-i+1} are (i-1) x (i-1) and (n-i+1) x (n-i+1) matrices respectively such that

$$AV = VA'. \tag{31}$$

Then we may write

$$D^{-1/2} V^T AVD^{-1/2} = D^{-1/2} V^T VA^1 D^{-1/2} = D^{1/2} A'D^{-1/2}. \tag{32}$$

Now $D^{1/2} A' D^{-1/2}$ is of the form of equation (30) and $VD^{-1/2}$ is an orthogonal matrix. Hence we can reduce the problem of calculating eigenvalues and eigenvectors of an nxn matrix to one of calculating eigenvalues and eigenvectors of an (i-1) x (i-1) matrix and an (n-i+1) x (n-i+1) matrix.

Case (2) $X_j^T X_j \neq 0 \quad 1 \leq j \leq n.$

If we define X_{n+1} by equations (25) and (26) then $X_{n+1} = 0$ because the X_j $1 \leq j \leq n$ are non-zero vectors spanning the n dimensional space. We define a matrix

$$V = (X_1, X_2, \ldots, X_n). \tag{33}$$

Now $V^TV = D$ where D is a diagonal matrix with positive elements on the diagonal. Hence D has a real square root $D^{1/2}$ and $D^{-1/2}$ exists, and $VD^{-1/2}$ is an orthogonal matrix. Hence

$$AV = VA' \tag{34}$$

where

$$A' = \begin{pmatrix} & & & \Large 0 \\ \alpha_1 & \beta_1 & & \\ 1 & \alpha_2 & \beta_2 & \\ 0 & 1 & \alpha_3 & \beta_3 \\ \Large 0 & & 1 & \end{pmatrix} . \qquad (35)$$

Therefore

$$D^{-1/2} V^T AVD^{-1/2} = D^{-1/2} V^T VA' D^{-1/2} = D^{1/2} A' D^{-1/2} . \qquad (36)$$

The right hand side of equation (36) is in tridiagonal form, and $VD^{-1/2}$ is an orthogonal matrix which transforms A to tridiagonal form.

The process of deflating a matrix is one essentially of isolating and eliminating from consideration one eigenvalue and a corresponding eigenvector from an nxn symmetric matrix A to obtain a matrix A' of the form

$$A' = \left(\begin{array}{c:c} \epsilon & 0 \\ \hdashline 0 & A_{n-1} \end{array}\right) \qquad (37)$$

where ϵ is an eigenvalue of A and A_{n-1} is an (n-1) x (n-1) matrix. Let ϵ, X be an eigenvalue and corresponding eigenvector, such that $X^TX = 1$ and let V be an nx(n-1) rectangular matrix such that $X^TV = 0$. Such matrices V can always be constructed. Consider the matrix $S = (X \vdots V)$. Then $S^TS = I$ and

$$A' = S^T AS = \left(\begin{array}{c} X^T \\ \hline V^T \end{array}\right) \quad A \quad \left(\begin{array}{c|c} X & V \end{array}\right)$$

$$= \left(\begin{array}{c} X^T \\ \hline V^T \end{array}\right) \quad \left(\begin{array}{c|c} eX & AV \end{array}\right)$$

$$= \left(\begin{array}{c|c} \epsilon & X^T AV \\ \hline \epsilon V^T X & V^T AV \end{array}\right) = \left(\begin{array}{c|c} \epsilon & 0 \\ \hline 0 & A_{n-1} \end{array}\right), \qquad (38)$$

which gives us the required reduction.

2.3 Sturm Sequence Method

This method is based upon the well known theorem that if A is a symmetric matrix and A_i is the i x i matrix formed from the upper left i^{th} order minor of A, then the eigenvalues of A_i are distinct from those of A_{i+1} and properly separate those of A_{i+1}. If $f_i(\epsilon)$ is the characteristic polynomial of A_i, namely $f_i(\epsilon) = \text{DET}(\epsilon I - A_i)$, $f_o(\epsilon)$ is defined to be one, and $f_n(\epsilon) \neq 0$, then the number of eigenvalues of A greater than ϵ is the number of sign changes in the sequence $f_o(\epsilon)$, $f_1(\epsilon)$, $f_2(\epsilon)$, ..., $f_n(\epsilon)$.

If A is in tridiagonal form, namely

$$A = \begin{pmatrix} \alpha_1 & \beta_1 & \bigcirc \\ \beta_1 & \alpha_2 & \beta_2 \\ \bigcirc & \beta_2 & \alpha_3 \end{pmatrix} \qquad (39)$$

the $f_i(\epsilon)$ may be given by the following formulas

$$f_o(\epsilon) = 1$$

$$f_1(\epsilon) = \epsilon - \alpha_1$$

$$f_i(\epsilon) = (\epsilon - \alpha_i)\, f_{i-1}(\epsilon) - \beta_{i-1}^2\, f_{i-2}(\epsilon) \qquad i > 2 \tag{40}$$

If some $\beta_i = 0$ the matrix decomposes into a direct sum of matrices, each of which may be considered separately. This must happen if A has multiple eigenvalues. Therefore, we may assume without loss of generality that all $\beta_i \neq 0$ and hence all eigenvalues are distinct.

The formulas in equation (4) may be used to determine the number of sign changes in the sequence $f_o(\epsilon)$, $f_1(\epsilon)$, ..., $f_n(\epsilon)$ and thereby determine intervals of any desired size in which each of the distinct eigenvalues may be found. It may be noted that the vanishing of an intermediate $f_j(\epsilon)$ $(1 \leq j < n)$ does not affect the number of sign changes in the above sequence because if $f_j(\epsilon) = 0$, then by equations (40) $f_{j-1}(\epsilon) \neq 0$, $f_{j+1}(\epsilon) \neq 0$ and $f_{j-1}(\epsilon)$, $f_{j+1}(\epsilon)$ must have opposite signs.

Once the eigenvalues ϵ have been found the corresponding eigenvectors $X^T = (x_1, x_2, \ldots, x_n)$ may be found by one of several straight forward methods. The simplest is that of back substitution. Take $x_n = 1$. Then

$$\begin{aligned} x_{n-1} &= (\epsilon - \alpha_n)/\, \beta_{n-1} \\ x_{n-2} &= ((\epsilon - \alpha_{n-1})\, x_{n-1} - \beta_{n-1})/\, \beta_{n-2} \\ x_{j-1} &= ((\epsilon - \alpha_j)\, x_j - \beta_j\, x_{j+1})/\, \beta_{j-1} \quad . \end{aligned} \tag{41}$$

However, this method leads to numerical instabilities should any of the β_i be small in absolute value.

Another method, referred to as the method of orthogonal factorization, is the following. Consider a matrix B

$$B = A - \epsilon I = \begin{pmatrix} u_1 & \beta_1 & \bigcirc \\ \beta_1 & u_2 & \beta_2 \\ \bigcirc & \beta_2 & u_3 \end{pmatrix} . \qquad (42)$$

β is singular so DET (B) $= 0$. Consider the following sequence of matrices $B^{(k)}$ $1 \leq k \leq n - 1$, defined as follows:

$$B^{(k)} = J^{(k)}_{k,k+1} B^{(k-1)} , \quad B^{(0)} = B. \qquad (43)$$

If we define the $c^{(1)}$, $s^{(1)}$ of $J^{(1)}_{12}$ as follows

$$c^{(1)} = u_1/(u_1^2 + \beta_1^2)^{1/2}$$

$$s^{(1)} = \beta_1/(u_1^2 + \beta_1^2)^{1/2} , \qquad (44)$$

then $B^{(1)}$ takes the form

$$B^{(1)} = \begin{pmatrix} d_1 & e_1 & f_1 & 0 & \bigcirc \\ 0 & g_2 & h_2 & 0 & \\ 0 & \beta_2 & u_3 & \beta_3 & \\ \bigcirc & & \beta_3 & & \end{pmatrix} \qquad (45)$$

where

$$\begin{aligned} d_1 &= (u_1^2 + \beta_1^2)^{1/2} \\ e_1 &= \beta_1 c^{(1)} + u_2 s^{(1)} \\ f_1 &= \beta_2 s^{(1)} \\ g_2 &= \beta_1 s^{(1)} - u_2 c^{(1)} \\ h_2 &= - \beta_2 c^{(1)} \quad . \end{aligned} \tag{46}$$

We define by induction

$$\begin{aligned} c^{(k)} &= g_k/(g_k^2 + \beta_k^2)^{1/2} \\ s^{(k)} &= \beta_k/(g_k^2 + \beta_k^2)^{1/2} \quad . \end{aligned} \tag{47}$$

Then the matrix $B^{(k)}$ appears thus:

$$B^{(k)} = \begin{pmatrix} d_1 & e_1 & f_1 & & & \bigcirc & \\ 0 & d_2 & e_2 & f_2 & & & \\ & & & d_k & e_k & f_k & \\ & & & 0 & g_{k+1} & h_{k+1} & \\ \bigcirc & & & & \beta_{k+1} & u_{k+2} & \beta_{k+2} \end{pmatrix} \tag{48}$$

where

$$\begin{aligned} d_k &= (g_k^2 + \beta_k^2)^{1/2} \\ e_k &= h_k\, c^{(k)} + u_{k+1}\, s^{(k)} \\ g_{k+1} &= h_k\, s^{(k)} - u_{k+1}\, c^{(k)} \\ h_{k+1} &= -\beta_{k+1}\, c^{(k)} \\ g_n &= d_n \quad . \end{aligned} \tag{49}$$

Now let $U = J^{(n-1)}_{n-1,n}\ J^{(n-2)}_{n-2,n-1}\ \ldots\ J^{(2)}_{2,3}\ J^{(1)}_{1,2}$. Then UB is upper triangular with only the first three upper diagonals having elements different from zero. If DET (B) = 0 then DET (UB) = 0 and hence $d_k = 0$ for some k. If $d_k = 0$ for $k \neq n$ then by equations (49) $\beta_k = 0$, contrary to assumption. So $d_n = 0$. If E^T is the vector (0, 0, ..., 0, 1) then $E^T\, UB = 0$ or $BU^T\, E = 0$ or $(A - \epsilon I)\, U^T E = 0$. Hence $AU^T\, E = \epsilon U^T\, E$ and $U^T E$ is an eigenvector of A corresponding to the eigenvalue ϵ.

A third method, referred to as the method of inverse iterations, is the following. Let ϵ_j and U_j be the corresponding eigenvalues and eigenvectors of A and let $\overline{\epsilon}_i$ be the calculated approximation to ϵ_i. Let $X^{(o)}$ be an arbitrary vector which we expand as follows:

$$X^{(0)} = \sum_{k=1}^{n} c_k\, U_k. \tag{50}$$

We assume that $c_i \neq 0$ and $X^{(0)T}\, X^{(0)} = 1$. Consider a sequence of vectors $X^{(j)}$ defined by

$$X^{(j-1)} = (A - \overline{\epsilon}_i)\, X^{(j)} \qquad j \geq 1. \tag{51}$$

If we set j = 1 in equation (51) we see that

$$X^{(1)} = \sum_{k=1}^{n} \frac{c_k U_k}{(\epsilon_k \bar{\epsilon}_i)} = \frac{c_i U_i}{(\epsilon_i - \bar{\epsilon}_i)} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{c_k U_k}{(\epsilon_k - \bar{\epsilon}_i)} \tag{52}$$

satisfies equation (51). More generally

$$X^{(j)} = \frac{c_i U_i}{(\epsilon_i \bar{\epsilon}_i)^j} + \sum_{\substack{k=1 \\ k \neq i}}^{n} \frac{c_k U_k}{(\epsilon_k - \bar{\epsilon}_i)^j} \;. \tag{53}$$

If $|\epsilon_i - \bar{\epsilon}_i|$ is small compared to $|\epsilon_k - \bar{\epsilon}_i|$ $k \neq i$, (which would be expected) then the $X^{(j)}$, if normalized, would converge to U_i as $j \longrightarrow \infty$.

2.4 Orthogonal Reduction Methods

This is a class of methods which constructs an orghogonal matrix U for a given real symmetric nxn matrix A such that

$$U^T AU = D \quad \text{(a diagonal matrix)} \tag{54}$$

by obtaining U as a limit

$$\lim_{k \longrightarrow \infty} U^{(k)} = U \tag{55}$$

where $U^{(k)}$ is usually a product of elementary orthogonal matrices U_j

$$U^{(k)} = U_1 U_2 U_3 \cdots U_k \;. \tag{56}$$

The columns of U will be the column eigenvectors with the corresponding diagonal element of D being the corresponding eigenvalue.

2.4.1 Jacobi Methods

In these methods the elementary orthogonal matrices are the Jacobi matrices J_{ij}. Let $A^{(o)} = A$,

$$A^{(k)} = J^{(k)}_{i_k j_k} \; A^{(k-1)} \; J^{(k)}_{i_k j_k} \tag{57}$$

$c^{(k)}$, $s^{(k)}$ be the numbers associated with the $J^{(k)}_{i_k j_k}$, and $\alpha^{(k)}_{pq}$ be the elements of $A^{(k)}$. The elements of $A^{(k)}$ are related to those of $A^{(k-1)}$ as follows:

$$\begin{aligned}
\alpha^{(k)}_{i_k i_k} &= c^{(k)} \left(c^{(k)} \alpha^{(k-1)}_{i_k i_k} + s^{(k)} \alpha^{(k-1)}_{i_k j_k}\right) \\
&\quad + s^{(k)} \left(c^{(k)} \alpha^{(k-1)}_{i_k j_k} + s^{(k)} \alpha^{(k-1)}_{j_k j_k}\right) \\
\alpha^{(k)}_{j_k j_k} &= s^{(k)} \left(s^{(k)} \alpha^{(k-1)}_{i_k i_k} - c^{(k)} \alpha^{(k-1)}_{i_k j_k}\right) \\
&\quad - c^{(k)} \left(s^{(k)} \alpha^{(k-1)}_{i_k j_k} - c^{(k)} \alpha^{(k-1)}_{j_k j_k}\right) \\
\alpha^{(k)}_{i_k j_k} &= s^{(k)} \left(c^{(k)} \alpha^{(k-1)}_{i_k i_k} + s^{(k)} \alpha^{(k-1)}_{i_k j_k}\right) \\
&\quad - c^{(k)} \left(c^{(k)} \alpha^{(k-1)}_{i_k j_k} + s^{(k)} \alpha^{(k-1)}_{j_k j_k}\right) \qquad (58) \\
&\left.\begin{aligned}
\alpha^{(k)}_{i_k q} &= c^{(k)} \alpha^{(k-1)}_{i_k q} + s^{(k)} \alpha^{(k-1)}_{j_k q} \\
\alpha^{(k)}_{j_k q} &= s^{(k)} \alpha^{(k-1)}_{i_k q} - c^{(k)} \alpha^{(k-1)}_{j_k q}
\end{aligned}\right\} \quad q \neq i_k,\ j_k \\
&\left.\begin{aligned}
\alpha^{(k)}_{p i_k} &= c^{(k)} \alpha^{(k-1)}_{p i_k} + s^{(k)} \alpha^{(k-1)}_{p j_k} \\
\alpha^{(k)}_{p j_k} &= s^{(k)} \alpha^{(k-1)}_{p i_k} - c^{(k)} \alpha_{p j_k}
\end{aligned}\right\} \quad p \neq i_k,\ j_k \\
\alpha^{(k)}_{pq} &= \alpha^{(k-1)}_{pq} \qquad p,\ q \neq i_k,\ j_k \ .
\end{aligned}$$

From equation (58) we can easily verify

$$
\begin{aligned}
&(\alpha^{(k)}_{i_k i_k})^2 + 2\,(\alpha^{(k)}_{i_k j_k})^2 + (\alpha^{(k)}_{j_k j_k})^2 \\
&= (\alpha^{(k-1)}_{i_k i_k})^2 + 2\,(\alpha^{(k-1)}_{i_k j_k})^2 + (\alpha^{(k-1)}_{j_k j_k})^2 \\
&(\alpha^{(k)}_{i_k q})^2 + (\alpha^{(k)}_{j_k q})^2 = (\alpha^{(k-1)}_{i_k q})^2 + (\alpha^{(k-1)}_{j_k q})^2 \qquad q \neq i_k,\ j_k \\
&(\alpha^{(k)}_{p i_k})^2 + (\alpha^{(k)}_{p j_k})^2 = (\alpha^{(k-1)}_{p i_k})^2 + (\alpha^{(k-1)}_{p j_k})^2 \qquad p \neq i_k,\ j_k\ .
\end{aligned}
\tag{59}
$$

If we set

$$
\frac{c^{(k)}}{s^{(k)}} = \frac{\alpha^{(k-1)}_{i_k i_k} + \alpha^{(k-1)}_{j_k j_k} \pm \sqrt{\left(\alpha^{(k-1)}_{i_k i_k} - \alpha^{(k-1)}_{j_k j_k}\right)^2 + 4\left(\alpha^{(k-1)}_{i_k j_k}\right)^2}}{2\,\alpha^{(k-1)}_{i_k j_k}} \qquad (\alpha^{(k-1)}_{i_k j_k} \neq 0)
\tag{60}
$$

then it is easily seen that $\alpha^{(k)}_{i_k j_k} = 0.$

Now we define

$$
\begin{aligned}
d^{(k)} &= \sum_{p=1}^{n} (\alpha^{(k)}_{pp})^2 \\
u^{(k)} &= \sum_{p=1}^{n} \sum_{\substack{q=1 \\ q\neq p}}^{n} (\alpha^{(k)}_{pq})^2 \\
f^{(k)} &= \sum_{p=1}^{n} \sum_{q=1}^{n} (\alpha^{(k)}_{pq})^2
\end{aligned}
\tag{61}
$$

then $f^{(k)} = f^{(k-1)} = \dots = f^{(o)} = f$, $f = d^{(k)} + u^{(k)}$. If $c^{(k)}/s^{(k)}$ is chosen by equation (60), then

$$d^{(k)} = d^{(k-1)} + 2\,(\alpha^{(k-1)}_{i_k j_k})^2$$
$$u^{(k)} = u^{(k-1)} - 2\,(\alpha^{(k-1)}_{i_k j_k})^2 \qquad (62)$$

by the fact that $\alpha^{(k)}_{i_k j_k} = 0$ and equations (59). If i_k, j_k are chosen so that

$$(\alpha^{(k-1)}_{i_k j_k})^2 \geq \frac{u^{(k-1)}}{n(n-1)} \qquad (63)$$

then

$$u^{(k)} \leq \left[1 - \frac{2}{n(n-1)}\right] u^{(k-1)} \leq \left[1 - \frac{2}{n(n-1)}\right]^k u^{(0)}. \qquad (64)$$

Hence

$$\lim_{k \to \infty} u^{(k)} = 0$$
$$\lim_{k \to \infty} A^{(k)} = \text{diagonal matrix} \qquad (65)$$
$$\lim_{k \to \infty} U^{(k)} = \lim_{k \to \infty} J^{(1)}_{i,j}\, J^{(2)}_{i_2 j_2} \cdots J^{(k)}_{i_k j_k} = U,$$

an orthogonal matrix of eigenvectors of A.

There is a considerable amount of freedom in selecting i_k, j_k at each step. We may, for example, choose i_k, j_k so that $\alpha^{(k-1)}_{i_k j_k}$ has the largest absolute value of all off-diagonal elements. We may also choose i_k, j_k sequentially as follows:

$$i_k = i_{k-1};\quad j_k = j_{k-1} + 1 \quad \text{for } i_{k-1} < n-1 \text{ and } j_{k-1} < n,$$
$$i_k = i_{k-1} + 1;\quad j_k = i_k + 1 \quad \text{for } i_{k-1} < n-1 \text{ and } j_{k-1} = n, \qquad (66)$$
$$i_k = 1,\ j_k = 2 \quad \text{for } i_{k-1} = n-1 \text{ and } j_{k-1} = n;$$

or

$$\begin{aligned}
j_k &= j_{k-1},\ i_k = i_{k-1} + 1 \quad \text{for } j_{k-1} < n \text{ and } 1_{k-1} < j_{k-1} - 1 \\
j_k &= j_{k-1} + 1,\ i_k = 1 \quad \text{for } j_{k-1} < n \text{ and } i_{k-1} = j_{k-1} - 1 \qquad (67) \\
i_k &= 1,\ j_k = 2 \quad \text{for } j_{k-1} = n \text{ and } i_{k-1} = j_{k-1} - 1 \ .
\end{aligned}$$

We obtain convergence for selection schemes equations (66) and (67) provided, if we set $c^{(k)} = \cos a^{(k)}$, $s^{(k)} = \sin a^{(k)}$, the angles $a^{(k)}$ all lie in some closed interval contained in the open interval $(-\Pi/2, \Pi/2)$.

2.4.2 La Budde-Kaiser Methods

In these methods Householder matrices will be used as elementary orthogonal matrices to construct the matrix U which diagonalizes A.

In the La Budde methods the iteration consists of two parts: one

$$H_1^{(k)}\ A^{(k-1)}\ H_1^{(k)} = B^{(k)} \qquad (68)$$

and, two

$$J_{12}^{(k)}\ B^{(k)}\ J_{12}^{(k)} = A^{(k)}\ . \qquad (69)$$

It is convenient to partition the matrices $A^{(k)}$ and $B^{(k)}$ as follows.

$$A^{(k)} = \left(\begin{array}{c|c} \alpha_{11}^{(k)} & S^{(k)T} \\ \hline S^{(k)} & A_1^{(k)} \end{array}\right) \qquad (70)$$

$$S^{(k)T} = (\alpha_{12}^{(k)} \ \vdots \ R^{(k)T})$$

$$A_1^{(k)} = \left(\begin{array}{c|c} \alpha_{22}^{(k)} & T^{(k)T} \\ \hline T^{(k)} & A_2^{(k)} \end{array}\right)$$

$$B^{(k)} = \left(\begin{array}{c|c} \beta_{11}^{(k)} & X^{(k)T} \\ \hline X^{(k)} & B_1^{(k)} \end{array} \right)$$

$$X^{(k)T} = (\beta_{12}^{(k)} \;\vdots\; Z^{(k)}$$

$$B_1^{(k)} = \left(\begin{array}{c|c} \beta_{22}^{(k)} & Y^{(k)T} \\ \hline Y^{(k)} & B_2^{(k)} \end{array} \right) .$$

Here $A_1^{(k)}$, $B_1^{(k)}$ are (n-1) x (n-1) matrices, $A_2^{(k)}$, $B_2^{(k)}$ are (n-2) x (n-2) matrices, $S^{(k)}$, $X^{(k)}$ are (n-1) dimensional vectors, and $R^{(k)}$, $T^{(k)}$, $Z^{(k)}$, $Y^{(k)}$ are (n-2) dimensional vectors.

If F is a matrix or vector, we define $||F||$ to be the square root of the sum of the squares of all of the elements of F. We now define the matrices $H_1^{(k)}$, $J_{12}^{(k)}$ used in the iteration. Let $H_1^{(k)} = I - 2W_i^{(k)} W_i^{(k)T}$ where $W_1^{(k)T} = (0, w_2^{(k)}, \ldots, w_n^{(k)})$. If we set

$$w_2^{(k)} = \left[1/2 \; (1 + |\alpha_{12}^{(k-1)}| \; / \; ||S^{(k)}|| \right]^{1/2}$$

$$w_j^{(k)} = \operatorname{sgn} (\alpha_{12}^{(k-1)}) \; \alpha_{1j}^{(k-1)} / \; (2w_2^{(k)} \; ||S^{(k)}||) \quad 3 \leq j \leq n \tag{71}$$

$$\begin{aligned}
\beta_{11}^{(k)} &= \alpha_{11}^{(k-1)} \\
|\beta_{12}^{(k)}|^2 &= ||S^{(k-1)}||^2 = (\alpha_{12}^{(k-1)})^2 + ||R^{(k-1)}||^2 \\
Z^{(k)} &= 0 \\
\beta_{22}^{(k)} &= V^{(k-1)T} A_1^{(k-1)} V^{(k-1)} \\
||Y^{(k)}||^2 &= V^{(k-1)T} A_1^{(k-1)} (I - V^{(k-1)} V^{(k-1)T}) A_1^{(k-1)} V^{(k-1)} \\
||B_1^{(k)}|| &= ||A_1^{(k-1)}||
\end{aligned} \tag{72}$$

where $V^{(k)} = S^{(k)}/||S^{(k)}||$. The effect of $J_{12}^{(k)}$ on $B^{(k)}$ is as follows. Let $c^{(k)}$, $s^{(k)}$ be the quantities associated with $J_{12}^{(k)}$. Then we have

$$\alpha_{11}^{(k)} = \beta_{11}^{(k)} + \Delta\,\alpha_{11}^{(k)} = \alpha_{11}^{(k-1)} + \Delta\,\alpha_{11}^{(k)} \tag{73}$$

where

$$\Delta\,\alpha_{11}^{(k)} = (s^{(k)})^2\,(\beta_{22}^{(k)} - \alpha_{11}^{(k-1)} + 2\,\frac{c^{(k)}}{s^{(k)}}\;\beta_{12}^{(k)}) \tag{74}$$

and

$$\alpha_{12}^{(k)} = \beta_{12}^{(k)} + \Delta\,\beta_{12}^{(k)} \tag{75}$$

where

$$\Delta\,\beta_{12}^{(k)} = (s^{(k)})^2\left(-2\left(\frac{c^{(k)}}{s^{(k)}}\right)^2 \beta_{12}^{(k)} + \frac{c^{(k)}}{s^{(k)}}\;(\alpha_{11}^{(k-1)} - \beta_{22}^{(k)})\right). \tag{76}$$

Finally

$$\begin{aligned}
||R^{(k)}|| &= |s^{(k)}|\;||Y^{(k)}|| \\
||T^{(k)}|| &= |c^{(k)}|\;||Y^{(k)}|| \\
A_2^{(k)} &= B_2^{(k)}\ .
\end{aligned} \tag{77}$$

It can easily be shown from equations (72 to 77) that if $d^{(k)}$ is any sequence of numbers satisfying the following: (1) $d^{(k)} > 0$, (2) $|d^{(k)}|$ are bounded away from zero, (3) $|d^{(k)}|$ are bounded away from ∞, and

$$\frac{c^{(k)}}{s^{(k)}} = \frac{\alpha_{11}^{(k-1)} - \beta_{22}^{(k)} + d^{(k)}}{2\,\beta_{12}^{(k)}} \tag{78}$$

then

$$\lim_{k \longrightarrow \infty} |\beta_{12}^{(k)}| = \lim_{k \longrightarrow \infty} ||S^{(k)}|| = 0 \tag{79}$$

$$\lim_{k \longrightarrow \infty} A^{(k)} = \left(\begin{array}{c|c} \epsilon_1 & 0 \\ \hline 0 & A_1 \end{array} \right)$$

where A_1 is an (n-1) x (n-1) matrix and ϵ_1 is an eigenvalue of A. Equation (79) is true if any one, two, or all of the conditions (1) - (3) are replaced by the corresponding conditions (1)' - (3)' : (1)' $d^{(k)} < 0$, (2)' $|d^{(k)}|$ approaches zero no faster than $|\beta_{12}^{(k)}|^2$, (3)' $|d^{(k)}|^{-1}$ approaches zero no faster than $|\beta_{12}^{(k)}|^2$.

It can also be shown by equations (72 to 77) that if

$$0 \leq \left|\frac{c^{(k)}}{s^{(k)}}\right| \leq \left|\frac{\alpha_{11}^{(k-1)} - \beta_{22}^{(k)}}{2\,\beta_{12}^{(k)}}\right| \tag{80}$$

$$\operatorname{sgn} \frac{c^{(k)}}{s^{(k)}} = \operatorname{sgn} \frac{\alpha_{11}^{(k-1)} - \beta_{22}^{(k)}}{2\,\beta_{12}^{(k)}}$$

then

$$\lim_{k \longrightarrow \infty} ||R^{(k)}|| = 0, \quad \lim_{k \longrightarrow \infty} ||T^{(k)}|| = 0 \tag{81}$$

$$\lim_{k \longrightarrow \infty} A^{(k)} = \left(\begin{array}{c|c} E_2 & 0 \\ \hline 0 & A_2 \end{array} \right)$$

where E_2 is a 2 x 2 matrix and A_2 is an (n-2) x (n-2) matrix.

Two special cases of interest will be noted here. One is the case where we define

$$d^{(k)} = \pm\sqrt{(\alpha_{11}^{(k-1)} - \beta_{22}^{(k)})^2 + 4\,(\beta_{12}^{(k)})^2} \quad . \qquad (82)$$

These choices of $d^{(k)}$ cause $\alpha_{12}^{(k)} = 0$ and maximize $|\Delta\, \alpha_{11}^{(k)}|$. The other case is the one in which we set $c^{(k)}/s^{(k)} = 0$. In this case the Jacobi transformation $J_{12}^{(k)}\, B^{(k)}\, J_{12}^{(k)}$ reduces to a simple row and column permutation which is exact and requires no computation.

The Kaiser iteration may be defined as follows

$$H_o^{(k)}\, A^{(k-1)}\, H_o^{(k)} = A^{(k)} \qquad (83)$$

where

$$\begin{aligned} H_o^{(k)} &= I - 2\, W_o^{(k)}\, W_o^{(k)T} \\ W_o^{(k)T} &= (w_1^{(k)}, w_2^{(k)}, \ldots, w_n^{(k)}) \quad . \end{aligned} \qquad (84)$$

Before proceeding with the k^th step we make sure that all elements $\alpha_{1j}^{(k-1)}$ $2 \leq j \leq n$ are non-negative. This can easily be done by appropriate row and column multiplications. We define

$$\begin{aligned} t^{(k-1)} &= \Big(\sum_{j=2}^{n} \alpha_{1j}^{(k-1)}\Big)/\sqrt{n-1} \\ u^{(k-1)} &= \Big(\sum_{j=2}^{n} \sum_{p=2}^{n} \alpha_{jp}^{(k-1)}\Big)/\,(n-1) \end{aligned} \qquad (85)$$

and we take

$$\begin{aligned} w_1^{(k)} &= c^{(k)} \\ w_2^{(k)} &= w_3^{(k)} = \ldots = w_n^{(k)} = s^{(k)} \quad . \end{aligned} \qquad (86)$$

Then we have

$$
\begin{aligned}
&(c^{(k)})^2 + (n-1)\,(s^{(k)})^2 = 1 \\
&\alpha_{11}^{(k)} = \alpha_{11}^{(k-1)}\,(1 - 2\,(c^{(k)})^2)^2 \\
&- 4\,\sqrt{n-1}\;\; c^{(k)}\, s^{(k)}\,(1 - 2\,(c^{(k)})^2\, t^{(k-1)} \\
&+ 4\,(n-1)\,(c^{(k)}\, s^{(k)})^2\, u^{(k-1)} \quad .
\end{aligned}
\tag{87}
$$

If we set

$$
\begin{aligned}
&\cos\; a^{(k)} = 1 - 2\,(c^{(k)})^2 \\
&\sin\; a^{(k)} = 2\,\sqrt{n-1}\;\; c^{(k)}\, s^{(k)}
\end{aligned}
\tag{88}
$$

then

$$(\cos a^{(k)})^2 + (\sin a^{(k)})^2 = 1 \tag{89}$$

and

$$
\begin{aligned}
\alpha_{11}^{(k)} &= (\cos a^{(k)})^2\;\alpha_{11}^{(k-1)} - 2\;\cos a^{(k)}\,\sin a^{(k)}\; t^{(k-1)} \\
&\quad + (\sin a^{(k)})^2 \quad {}^{(k-1)} \\
&= 1/2\,(\alpha_{11}^{(k-1}) + \quad {}^{(k-1)}) + 1/2\,\cos\,2a^{(k)}\,(\alpha_{11}^{(k-1)} - \quad {}^{(k-1)} \\
&\qquad\qquad - \sin\,2a^{(k)}\; t^{(k-1)} \quad .
\end{aligned}
\tag{90}
$$

If we choose

$$\tan\,2a^{(k)} = -2t^{(k-1)}/(\alpha_{11}^{(k-1)} - u^{(k-1)}) \tag{91}$$

then $\alpha_{11}^{(k)}$ assumes its maximum value as a function of $a^{(k)}$ and is

$$\alpha_{11}^{(k)} = 1/2\ (\alpha_{11}^{(k-1)} + u^{(k-1)}) \qquad (92)$$

$$+ \ 1/2 \left[(\alpha_{11}^{(k-1)} - u^{(k-1)}) + 4\ (t^{(k-1)})^{2}\right]^{1/2} .$$

This shows that $\alpha_{11}^{(k)} \geq \alpha_{11}^{(-1)}$ with equality holding only if $t^{(k-1)} = 0$. If the $a^{(k)}$ are chosen by equation (91) then equations (79) hold true and convergence is obtained.

2.4.3 The L-R Method

In this section we will assume that the nxn symmetric matrix A is positive definite and is in tridiagonal form, i.e.

$$A = \begin{pmatrix} \alpha_1 & \beta_1 & \bigcirc \\ \beta_1 & \alpha_2 & \beta_2 \\ \bigcirc & \beta_2 & \alpha_3 \end{pmatrix} . \qquad (93)$$

We may also assume, without loss of generality as before that all $\beta_i \neq 0$.

The basic iteration is as follows.

We factor $A^{(k-1)}$ into

$$A^{(k-1)} = L^{(k-1)T}\ L^{(k-1)} \qquad (94)$$

where

$$L^{(k-1)} = \begin{pmatrix} d_1^{(k-1)} & e_1^{(k-1)} & \bigcirc \\ & d_2^{(k-1)} & e_2^{(k-1)} \\ & & \\ \bigcirc & & \end{pmatrix} . \qquad (95)$$

$L^{(k-1)}$ is upper triangular with non-zero elements only on the main and first upper diagonals. We now form

$$A^{(k)} = L^{(k-1)} \; L^{(k-1)T} \tag{96}$$

$A^{(k)}$ is similar to $A^{(k-1)}$ i.e.

$$A^{(k)} = (L^{(k-1)T})^{-1} \quad A^{(k-1)} \; L^{(k-1)T} \quad . \tag{96}$$

Also, $A^{(k)}$ is in tridiagonal form.

The relations between the elements of $A^{(k-1)}$, $L^{(k-1)}$, and $A^{(k)}$ are as follows

$$\begin{aligned}
d_1^{(k-1)} &= \sqrt{\alpha_1^{(k-1}} \\
e_1^{(k-1)} &= \beta_1^{(k-1)}/d_1^{(k-1)} \\
d_j^{(k-1)} &= (\alpha_j^{(k-1)} - (e_{j-1}^{(k-1)})^2)^{1/2} && 2 \leq j \leq n \\
e_j^{(k-1)} &= \beta_j^{(k-1)}/d_j^{(k-1)} && 2 \leq j \leq n - 1 \\
\alpha_1^{(k)} &= (d_1^{(k-1)})^2 + (e_1^{(k-1)})^2 \\
&= \alpha_1^{(k-1)} + (\beta_1^{(k-1)})^2/ \alpha_1^{(k-1)} \\
\beta_1^{(k)} &= d_2^{(k-1)} \; e_1^{(k-1)} \\
&= \frac{\beta_1^{(k-1)}}{\sqrt{\alpha_1^{(k-1)}}} \sqrt{\alpha_2^{(k-1)} - (\beta_1^{(k-1)})^2/ \alpha_1^{(k-1)}} \\
\alpha_j^{(k)} &= (d_j^{(k-1)})^2 + (e_j^{(k-1)})^2 \\
&= \alpha_j^{(k-1)} - (\beta_{j-1}^{(k-1)}/ d_{j-1}^{(k-1)})^2 + (\beta_j^{(k-1)}/d_j^{(k-1)})^2 && 2 \leq j \leq n\text{-}1
\end{aligned} \tag{97}$$

$$\alpha_n^{(k)} = (d_n^{(k-1)})^2 = \alpha_n^{(k-1)} - (\beta_{n-1}^{(k-1)}/d_{n-1}^{(k-1)})^2$$

$$\beta_j^{(k)} = d_{j+1}^{(k-1)} \; e_j^{(k-1)}$$

$$= (\beta_j^{(k-1)}/d_j^{(k-1)})(\alpha_{j+1}^{(k-1)} - (\beta_j^{(k-1)}/d_j^{(k-1)})^2)^{1/2} \quad 2 \leq j \leq n-1 .$$

It can be shown from equations (97) that all $\beta_j^{(k)} \longrightarrow 0$ as $k \longrightarrow \infty$ and hence

$$\lim_{k \to \infty} A^{(k)} = \text{a diagonal matrix}$$

$$\lim_{k \to \infty} U^{(k)} = \lim_{k \to \infty} L^{(1)T} L^{(2)T} \ldots L^{(k)T} = U$$

where U is a matrix, the columns of which are the eigenvectors of A. U may be multiplied on the right by a suitable diagonal matrix D so that UD is orthogonal. In theory one may obtain an orthogonal matrix as a limit of a product of non-orthogonal matrices $L^{(j)T}$, but in practice it may be better to obtain the vectors by one of the methods of section 2.3.

We may speed up the iterations of equations (97) by eliminating the square roots from the process as follows

$$\begin{aligned}
(d_1^{(k-1)})^2 &= \alpha_1^{(k-1)} \\
(e_1^{(k-1)})^2 &= (\beta_1^{(k-1)}/d_1^{(k-1)})^2 \\
(d_j^{(k-1)})^2 &= \alpha_j^{(k-1)} - (e_{j-1}^{(k-1)})^2 \\
(e_j^{(k-1)})^2 &= (\beta_j^{(k-1)}/d_j^{(k-1)})^2
\end{aligned} \qquad (99)$$

$$\alpha_1^{(k)} = (d_1^{(k-1)})^2 + (e_1^{(k-1)})^2$$

$$(\beta_1^{(k)})^2 = (d_2^{(k-1)} \; e_1^{(k-1)})^2$$

$$\alpha_j^{(k)} = (d_j^{(k-1)})^2 + (e_j^{(k-1)})^2$$

$$(\beta_j^{(k)})^2 = (d_{j+1}^{(k-1)} \; e_j^{(k-1)})^2$$

$$\alpha_n^{(k)} = (d_n^{(k-1)})^2 \quad .$$

Here we ignore the signs of $\beta_i^{(k)}$ and store only $(\beta_i^{(k)})^2$. The limit of $A^{(k)}$ will be a diagonal matrix of eigenvalues of A and we may obtain the eigenvectors by one of the methods of section 2.3.

Convergence may also be accelerated by a series of origin shifts. Instead of the sequence $A^{(k)}$, we consider the modified sequence $\bar{A}^{(k)} = A^{(k)} - u_k I$ where the u_k are chosen to be close to the smallest eigenvalue of A or the smallest diagonal element of A. More complicated choices of u_k will insure cubic convergence to diagonal form.

2.4.4 The Q-R Method

The Q-R method is based upon the fact that any matrix A may be factored in a non-iterative fashion into a product A = UT where U is orthogonal and T is upper triangular. If A is symmetric and in tridiagonal form

$$A = \left(\begin{array}{cccc} \alpha_1 & \beta_1 & \bigcirc & \\ \beta_1 & \alpha_2 & \beta_2 & \\ & \beta_2 & \alpha_3 & \beta_3 \\ \bigcirc & & \beta_3 & \end{array} \right) \qquad (100)$$

then T is upper triangular with non-zero elements only on the main diagonal and the first two upper diagonals. If we form $A' = TU$, then A and A' are similar $(A' = U^T AU)$ and A' is also in tridiagonal form. We can assume as before, without loss of generality that all $\beta_j \neq 0$.

We will formulate the basic iteration as follows: $A^{(k)} = V^{(k)T} A^{(k-1)} V^{(k)}$ where $V^{(k)}$ is an orthogonal matrix defined by

$$V^{(k)} = J_{1,2}^{(k)(1)} J_{2,3}^{(k)(2)} \cdots\cdots J_{n-2,n-1}^{(k)(n-2)} J_{n-1,n}^{(k)(n-1)} . \tag{101}$$

Let $c^{(k)(j)}$, $s^{(k)(j)}$ be the quantities associated with the $J_{j,j+1}^{(k)(j)}$. It will be convenient to use the following notation:

$$\begin{aligned}
A^{(k-1)(0)} &= A^{(k-1)} \\
A^{(k-1)(j)} &= J_{j,j+1}^{(k)(j)} A^{(k-1)(j-1)} \\
A^{(k-1)(n-1)} &= B^{(k)(0)} \\
B^{(k)(j)} &= B^{(k)(j-1)} J_{j,j+1}^{(k)(j)} \\
B^{(k)(n-1)} &= A^{(k)(0)} = A^{(k)} .
\end{aligned} \tag{102}$$

The general forms of $A^{(k-1)(j)}$ and $B^{(k)(j)}$ are as follows

$$A^{(k-1)(j)} = \begin{pmatrix} h_1^{(k)} & \delta_1^{(k)} & \epsilon_1^{(k)} & & & & & & \bigcirc \\ 0 & h_2^{(k)} & \delta_2^{(k)} & \epsilon_2^{(k)} & & & & & \\ & & & \cdot & & & & & \\ & & & & h_j^{(k)} & \delta_j^{(k)} & \epsilon_j^{(k)} & & \\ & & & & 0 & g_{j+1}^{(k)} & u_{j+1}^{(k)} & 0 & 0 \\ \bigcirc & & & & 0 & \beta_{j+1}^{(k-1)} & \alpha_{j+2}^{(k-1)} & \beta_{j+1}^{(k-1)} & 0 \end{pmatrix} \qquad (103)$$

$$\beta^{(k)(j)} = \begin{pmatrix} \alpha_1^{(k)} & \beta_1^{(k)} & 0 & & & & \\ \beta_1^{(k)} & \alpha_2^{(k)} & \beta_2^{(k)} & & & & \bigcirc \\ & & & & \beta_{j-1}^{(k)} & 0 & \\ & & & \beta_{j-1}^{(k)} & \alpha_j^{(k)} & t_j^{(k)} & \epsilon_j^{(k)} \\ & & & 0 & \beta_j^{(k)} & v_{j+1}^{(k)} & \delta_{j+1}^{(k)} \\ \bigcirc & & & & 0 & 0 & h_{j+1}^{(k)} \end{pmatrix} \qquad (104)$$

We will now define the quantities $c^{(k)(j)}$ and $s^{(k)(j)}$ and the elements of $A^{(k-1)(j)}$ and $B^{(k)(j)}$.

$$\begin{aligned} s^{(k)(1)} &= \beta_1^{(k-1)}/((\alpha_1^{(k-1)})^2 + (\beta_1^{(k-1)})^2)^{1/2} \\ c^{(k)(1)} &= \alpha_1^{(k-1)}/((\alpha_1^{(k-1)}0^2 + (\beta_1^{(k-1)})^2)^{1/2} \ . \end{aligned} \tag{105}$$

From this it follows that

$$\begin{aligned} h_1^{(k)} &= ((\alpha_1^{(k-1)})^2 + (\beta_1^{(k-1)})^2)^{1/2} \\ \delta_1^{(k)} &= \beta_1^{(k-1)} (\alpha_1^{(k-1)} + \alpha_2^{(k-1)})/h_1^{(k)} \\ \epsilon_1^{(k)} &= s^{(k)(1)} \beta_2^{(k-1)} \\ g_2^{(k)} &= ((\beta_1^{(k-1)})^2 - \alpha_1^{(k-1)} \alpha_2^{(k-1)})/h_1^{(k)} \\ u_2^{(k)} &= - c^{(k)(1)} \beta_2^{(k-1)} \ . \end{aligned} \tag{106}$$

The remaining quantities are defined as follows:

$$\begin{aligned} s^{(k)(j)} &= \beta_j^{(k-1)}/((g_j^{(k)})^2 + (\beta_j^{(k-1)})^2)^{1/2} \\ c^{(k)(j)} &= g_j^{(k)}/((g_j^{(k)})^2 + (\beta_j^{(k-1)})^2)^{1/2} \ . \end{aligned} \tag{107}$$

Then it can be seen that

$$
\begin{aligned}
h_j^{(k)} &= ((g_j^{(k)})^2 + (\beta_j^{(k-1)})^2)^{1/2} , && 2 \leq j \leq n-1 \\
\delta_j^{(k)} &= \beta_j^{(k-1)}(-c^{(k)(j-1)}\, g_j^{(k)} + \alpha_{j+1}^{(k-1)})/h_j^{(k)}, && 2 \leq j \leq n-1 \\
\epsilon_j^{(k)} &= s^{(k)(j)}\, \beta_{j+1}^{(k-1)} && 2 \leq j \leq n-2 \\
g_{j+1}^{(k)} &= s^{(k)(j)}\, u_j^{(k)} - c^{(k)(j)}\, \alpha_{j+1}^{(k-1)} && 2 \leq j \leq n-1 \\
u_{j+1}^{(k)} &= -c^{(k)(j)}\, \beta_{j+1}^{(k-1} && 2 \leq j \leq n-2 \\
g_n^{(k)} &= h_n^{(k)} \quad .
\end{aligned}
\tag{108}
$$

For the elements of $B^{(k)(j)}$ we have the following relations

$$
\begin{aligned}
\alpha_1^{(k)} &= \alpha_1^{(k-1)} + (\alpha_1^{(k-1)} + \alpha_2^{(k-1)})(\beta_1^{(k-1)})^2/(h_1^{(k)})^2 \\
\beta_1^{(k)} &= s^{(k)(1)}\, h_2^{(k)} \\
t_1^{(k)} &= s^{(k)(1)}\, h_1^{(k)} -c^{(k)(1)}\, \delta_1^{(k)} \\
v_2^{(k)} &= -c^{(k)(1)}\, h_1^{(k)} \\
\alpha_j^{(k)} &= c^{(k)(j)}\, v_j^{(k)} + s^{(k)(j)}\, \delta_j^{(k)} && 2 \leq j \leq n-1 \\
\beta_j^{(k)} &= s^{(k)(j)}\, h_{j+1}^{(k)} && 2 \leq j \leq n-1 \\
t_j^{(k)} &= s^{(k)(j)}\, v_j^{(k)} -c^{(k)(j)}\, \delta_j^{(k)} && 2 \leq j \leq n-1 \\
v_{j+1}^{(k)} &= -c^{(k)(j)}\, h_{j+1}^{(k)} && 2 \leq j \leq n-1 \\
v_n^{(k)} &= \alpha_n^{(k)} \quad .
\end{aligned}
\tag{109}
$$

It can be shown from equations (105 and 106) that if A is positive definite then

$$\lim_{k \to \infty} A^{(k)} = (\text{diagonal matrix}), \tag{110}$$

and

$$\lim_{k \to \infty} U^{(k)} = \lim_{k \to} \infty \; V^{(1)} V^{(2)} \dots V^{(k)} = U \tag{111}$$

where U is an orthogonal matrix of column eigenvectors of A.

Equations (110 and 111) can be shown to hold true under the more general requirement that A need only have eigenvalues with distinct absolute values. Convergence can be accelerated by origin shifts, i.e. employing the modified sequence of matrices $\overline{A}^{(k)} = A^{(k)} - u_k I$ where u_k is taken to be close to the smallest eigenvalue (in absolute value) of A or the smallest, in absolute value, diagonal element of A.

2.5 Step by Step Methods

In these methods, an eigenvalue ϵ and a corresponding eigenvector X are obtained, one pair ϵ and X at a time. When an ϵ and X have been obtained, the matrix may be transformed by the method of deflation to a matrix of essentially one lower order. We will describe two such methods: the power method and the conjugate gradient method.

2.5. The Power Method

Let A be a real symmetric nxn matrix with eigenvalues ϵ_i so ordered that $|\epsilon_1| \geq |\epsilon_2| \geq \dots . \geq |\epsilon_n|$ and let U_i be the corresponding eigenvectors of A. Let X_o be a vector which we may write in terms of U_i as follows:

$$X_o = \sum_{i=1}^{n} v_i U_i . \tag{112}$$

We further define

$$X_k = A^k X_o = AX_{k-1} \quad . \tag{113}$$

Then

$$X_k = \sum_{i=1}^{n} v_i \, \epsilon_i^K \, U_i \quad .$$

We wish to study the behavior of X_k as $k \longrightarrow \infty$. We consider three cases.

Case 1 (all eigenvalues having distinct absolute values).

Here we have $|\epsilon_1|>|\epsilon_2|>\ldots\ldots>|\epsilon_n|$.

We now assume that $v_1 \neq 0$. Then X_k can be written:

$$X_k = \epsilon_1^k \left(v_1 U_1 + \sum_{i=2}^{n} v_i \left(\frac{\epsilon_i}{\epsilon_1}\right)^k U_i\right) \quad . \tag{115}$$

In this case $(\epsilon_i/\epsilon_1)^k$ approaches zero as $k \longrightarrow \infty$ for $i \neq 1$, so $X_k/||X_k||$ approaches the eigenvector U_1. When U_1 has been computed to sufficient accuracy, we may set $\epsilon_1 = U_1^T AU_1$, deflate the matrix A, and continue.

Case 2 (multiple eigenvalues, distinct eigenvalues having distinct absolute values).

Here it is sufficient to consider the case where $\epsilon_1 = \epsilon_2 = \ldots \epsilon_p$; $|\epsilon_p|>|\epsilon_{p+1}|>\ldots>|\epsilon_n|$. We now assume that at least one $v_i \neq 0$ for $1 \leq i \leq p$. Now any linear combination of the eigenvectors $U_1, \ldots, U_p$ will also be an eigenvector of A corresponding to the eigenvalue ϵ_1. We may write X_k as follows:

$$X_k = \epsilon_1^k \left(U + \sum_{i=p+1}^{n} v_i \left(\frac{\epsilon_i}{\epsilon_1}\right)^k U_i\right) \tag{116}$$

where

$$U = \sum_{i=1}^{k} v_i U_i \tag{117}$$

and U is an eigenvector of A corresponding to the eigenvalue ϵ_i. In this case $(\epsilon_i/\epsilon_1)^k$ approaches zero as $k \longrightarrow \infty$ for $i > p$, so $X_k/||X_k||$ approaches the eigenvector U. When U has been computed to sufficient accuracy, we may proceed as in Case 1.

Case 3 (unequal eigenvalues with equal absolute values).

Here it is sufficient to consider the case where $\epsilon_1 = -\epsilon_2$, $|\epsilon_1| = |\epsilon_2| > |\epsilon_3| > \ldots >| \; |\epsilon_n|$. We assume that $v_1 \neq 0$ and $v_2 \neq 0$. Then X_k can be written

$$X_k = \epsilon_1^k \, (v_1 U_1 + (-1)^k \, v_2 U_2 + \sum_{i=3}^{n} v_i \, (\frac{\epsilon_i}{\epsilon_1})^k \; U_i) \qquad (118)$$

or

$$\begin{aligned} X_{2\ell} &= \epsilon_1^{2\ell} \, (v_1 U_1 + v_2 U_2 + \sum_{i=3}^{n} v_i \, (\frac{\epsilon_i}{\epsilon_1})^{2\ell} \; U_i) \\ X_{2\ell+1} &= \epsilon_1^{2\ell+1} \, (v_1 U_1 - v_2 U_2 = \sum_{i=3}^{n} v_i \, (\frac{\epsilon_i}{\epsilon_1})^{2\ell+1} \; U_i) \; . \end{aligned} \qquad (119)$$

The first of equations (119), for example, may be used to determine ϵ_1^2 by the methods of cases 1 and 2. To determine the eigenvectors U_1 and U_2, we take two sequences

$$\begin{aligned} Y_{2\ell} &= X_{2\ell} + \epsilon_1 \, X_{2\ell-1} \\ Z_{2\ell} &= Z_{2\ell} - \epsilon_1 \, X_{2\ell-1} \; . \end{aligned} \qquad (120)$$

Then

$$\begin{aligned} Y_{2\ell} &= \epsilon_1^{2\ell} \, (2 v_2 U_2 + \sum_{i=3}^{n} v_i \, (\frac{\epsilon_i}{\epsilon_1} + \epsilon_1) \, (\frac{\epsilon_i}{\epsilon_1})^{2\ell-1} \; U_i) \\ Z_{2\ell} &= \epsilon_1^{2\ell} \, (2 v_2 U_2 = \sum_{i=3}^{n} v_i \, (\frac{\epsilon_i}{\epsilon_1} - \epsilon_1) \, (\frac{\epsilon_i}{\epsilon_1})^{2\ell-1} \; U_i) \end{aligned} \qquad (121)$$

and as $\ell \longrightarrow \infty$ $Y_{2\ell}/||Y_{2\ell}||$ approaches U_1 and $Z_{2\ell}/||Z_{2\ell}||$ approaches U_2. We may then deflate the matrix twice using ϵ_1, U_1, $-\epsilon_1$, U_2 and continue as before.

In the course of the actual computation we may determine that we have cases 1 or 2 by noting that the iterates $X_k/||X_k||$ converge smoothly toward a limit vector U. Case 3 may be detected by noting that the successive iterates $X_k/||X_k||$ and $X_{k+1}/||X_{k+1}||$ oscillate between two limiting vectors and the special formulas of that case may be applied.

Convergence may be accelerated by means of origin shifts, i.e., employing a modified sequence of vectors $\bar{X}_k = (A - u_k I) X_{k-1}$. We may choose u_k, for example, to be near $(1/2)(|\epsilon_2| + |\epsilon_n|)$ or near the average of the absolute values of the two diagonal elements having the smallest absolute value and the second largest absolute value.

2.5.2 The Conjugate Gradient Method

We know that the largest and smallest eigenvalues of a real symmetric matrix A is given by the maximum and minimum respectively of the expression $(X^T AX)/(X^T X)$ as a function of X where X is a vector. The conjugate gradient method is a method for obtaining the maximum or minimum of that expression and may be described as follows.

Let X_o be an arbitrary vector such that $X_o^T X_o = 1$. Then we define a sequence of vectors X_k as follows. We assume that $X_{k-1}^T X_{k-1} = 1$ and define

$$\begin{aligned} t_{k-1} &= X_{k-1}^T AX_{k-1} \\ Z_K &= AX_{k-1} - t_{k-1} X_{k-1} = AX_{k-1} - X_{k-1} X_{k-1}^T AX_{k-1} \qquad (122) \\ g_k &= Z_k^T AX_{k-1} \;. \end{aligned}$$

We note that $||Z_k|| \leq 2||A||$ and $X_{k-1}^T Z_k = 0$ for all k. We will show that if either $Z_k = 0$ or $g_k = 0$ then X_{k-1} is an eigenvector of A. If $Z_k = 0$, then X_{k-1} is an eigenvector of A by the second of equations (122). If $g_k = 0$ then we have

$$\begin{aligned} 0 = g_k &= Z_k^T AX_{k-1} = (X_{k-1}^T A - X_{k-1}^T AX_{k-1} X_{k-1}^T) AX_{k-1} \\ &= X_{k-1}^T A(I - X_{k-1} X_{k-1}^T) AX_{k-1} \\ &= X_{k-1}^T A(I - X_{k-1} X_{k-1}^T)(I - X_{k-1} X_{k-1}^T) AX_{k-1} \qquad (123) \\ &= \left[(I - X_{k-1} X_{k-1}^T) AX_{k-1}\right]^T \left[(I - X_{k-1} X_{k-1}^T) AX_{k-1}\right] , \end{aligned}$$

since $I - X_{k-1} X_{k-1}^T$ is a projection onto the subspace of all vectors orthogonal to X_{k-1}. Equation (123) implies

$$(I - X_{k-1} X_{k-1}^T) AX_{k-1} = 0 \qquad (124)$$

which can only happen if X_{k-1} is an eigenvector of A. Thus we may only consider the case where $Z_k \neq 0$ and $g_k \neq 0$. We define

$$\begin{aligned} Y_k &= Z_k/||Z_k|| \\ h_k &= g_k/||Z_k|| = Y_k^T AX_{k-1} \\ u_k &= Y_k^T AY_k \qquad (125) \\ \bar{X}_k &= X_{k-1} + \alpha_k Y_k \\ X_k &= \bar{X}_k/||\bar{X}_k|| \end{aligned}$$

where α_k remains to be chosen. We note that $\bar{X}_k \neq 0$ for all choices of α_k. We now form t_k

$$
\begin{aligned}
t_k &= X_k^T AX_k = (\bar{X}_k^T A\bar{X}_k)/(\bar{X}_k^T \bar{X}_k) \\
&= \frac{t_{k-1} + 2h_k \alpha_k + u_k \alpha_k^2}{1 + \alpha_k^2} \\
&= t_{k-1} + \frac{2h_k \alpha_k + (u_k - t_{k-1})\alpha_k^2}{1 + \alpha_i^2} \\
&= t_{k-1} + \frac{1}{1 + (1/\alpha_k)^2} (2h_k/\alpha_k + (u_k - t_{k-1})) \ .
\end{aligned}
\tag{126}
$$

Let δ_k be any sequence of numbers and set

$$\alpha_k = \frac{2h_k}{t_{k-1} - u_k + \delta_k} \quad . \tag{127}$$

Then equation (126) becomes

$$t_k = t_{k-1} + \frac{4h_k^2 \ \delta_k}{4h_k^2 + (t_{k-1} - u_k + \delta_k)^2} \quad . \tag{128}$$

If the δ_k are now subjected to the following restrictions: (1) $\delta_k > 0$, (2) $|\delta_k|$ is bounded away from zero, (3) $|\delta_k|$ is bounded away from ∞, then the t_k form a monotonically increasing sequence bounded from above. Hence the t_k must converge to a limit, namely the maximum of $(X^TAX)/X^TX)$, and $t_k - t_{k-1}$ must approach zero as $k \longrightarrow \infty$. But by the conditions imposed on the δ_k and equation (128) h_k and hence g_k must approach zero as $k \longrightarrow \infty$. Hence the X_k approach an eigenvector of A and the t_k approach a corresponding eigenvalue. Once the eigenvalue and eigenvector pair have been computed to sufficient accuracy, the process of deflation may be used on A and the procedure may be continued.

We also obtain convergence if one, two, or all of the conditions (1) - (3) are replaced by the corresponding conditions (1)' $\delta_k < 0$, (2)' $|\delta_k| \longrightarrow 0$ no faster than h_k^2, (3)' $|1/\delta_k| \longrightarrow 0$ no faster than h_k^2. Condition (1)' will lead to the computation of the minimum of $(X^T AX)/X^T X)$. The choices of δ_k which maximize $|t_k - t_{k-1}|$ are given by

$$\delta_k = \pm \sqrt{(t_{k-1} - u_k)^2 + 4h_k^2} \qquad (129)$$

or

$$\alpha_k = \frac{u_k - t_{k-1} \pm \sqrt{(t_{k+1} - u_k)^2 + 4h_k^2}}{2h_k} \quad . \qquad (130)$$

3. Conclusions

Of the three classes of methods described in this report, the step-by-step methods appear to be the poorest for the following two reasons: (1) they are generally slow in convergence, relative to the other methods, even with judicious choices of origin shifts; (2) there is a cumulative loss of accuracy as one proceeds by deflation to calculate the eigenvalues and eigenvectors occurring later in the computation.

The orthogonal reduction methods (with the exception of the $L - L^T$ method) have the advantage that the eigenvectors are built up as a product of elementary orthogonal matrices, a process which is numerically stable and guarantees a set of eigenvector approximations which is reasonably orthogonal. In the $L - L^T$ method we obtain the orthogonal eigenvector matrix as a product of elementary non-orthogonal matrices, a procedure which is not recommended from the point of view of numerical stability or accuracy. However, the orthogonal reduction methods (with the exception of the $L - L^T$ method) are slow as compared with the Sturm sequence method and the square root free modification of the $L - L^T$ method, although they are faster than the step-by-step methods.

The fastest methods for obtaining the eigenvalues of a real symmetric matrix are the Sturm sequence method, applicable to general real symmetric matrices and the square root free modification of the $L - L^T$ method, applicable to positive definite real symmetric matrices. Of the two methods, the latter is faster when applicable. Of course, a real symmetric matrix can always be made positive definite by a suitable origin shift, but this can result in a loss of accuracy, particularly if the required origin shift is large.

Hence, for the calculation of the eigenvalues of real symmetric matrices the Sturm sequence method is recommended in the general case and the square root free modification of the $L - L^T$ method is recommended in the positive definite case. Once the eigenvalues have been obtained, the method of inverse iterations appears to be an effective and stable method for obtaining the eigenvectors. The reduction to tridiagonal form, prior to the application of the Sturm sequence or square root free modification of the $L - L^T$ method, may be most quickly and stably carried out by the Householder method.

ACKNOWLEDGEMENTS

The authors are indebted to Dr. B. C. Sawhney and Dr. M. J. Hanley of New York University for many helpful suggestions and discussions. They are also grateful to Mrs. Ida Kelly for her assistance in the preparation of the manuscript.

REFERENCES

1. Aitken, A. C.: The Evaluation of the Latent Roots and Latent Vectors of A Matrix. Proc. Roy. Soc. (Edin.) Sec. A, 1937, pp. 269-304.

2. Arnoldi, W. E.: The Principles of Minimized Iterations in the Solution of the Matrix Eigenvalue Problem. Quart. Appl. Math, vol. 9, 1951, pp. 17-29.

3. Bauer, F. L.: Sequential Reduction to Tridiagonal Form. J. Soc. Indus. Appl. Math., vol. 7, 1959, pp. 107-113.

4. Bauer, F. L., and Householder, A. S.: Moments and Characteristic Roots. Numer. Math., vol. 2, 1960, pp. 42-53.

5. Bellman, Richard: Introduction to Matrix Analysis. McGraw Hill Book Co., Inc., New York, 1960.

6. Bodewig, E.: Matrix Calculus. Interscience Publishers, Inc., New York, 1959.

7. Brooker, R. A., and Sumner, F. H.: The Method of Lanczos for Calculating the Characteristic Roots and Vectors of A Real Symmetric Matrix. Proc. Inst. Elect. Engrs.., B 103 Suppl., 1956, pp. 114-119.

8. Causey, R. L., and Gregory, R. T.: On Lanczos' Algorithm for Tridiagonalizing Matrices, SIAM Rev. vol. 3, 1961, pp. 322-328.

9. Faddeev, D. K., and Faddeeva, V. N.: Computational Methods of Linear Algebra. Freeman and Co., San Francisco, 1960.

10. Forsythe, G. E.: Contemporary State of Numerical Analysis; Surveys in Applied Mathematics V, John Wiley and Sons, 1958, pp. 3-42.

11. Forsythe, G. E., and Henrici, P.: The Cyclic Jacobi Method for Computing the Principal Values of a Complex Matrix. Trans. Amer. Math. Soc., vol. 94, 1960, pp. 1-23.

12. Francis, J. G. F.: The QR Transformations. Parts I and II, Computer Journal, vol. 4, 1961 and 1962, pp. 265-271, pp. 332-345.

13. Givens, R.: A Method of Computing Eigenvalues and Eigenvectors Suggested by Classical Results on Symmetric Matrices. Nat. Bur. Standards. Appl. Math, vol. 29, 1953, pp. 117-122.

14. Givens, W.: Numerical Computation of the Characteristic Values of A Real Symmetric Matrix. Oak Ridge Natl. Lab., Report ORNL - 1574, 1954.

15. Goldstine, H. H., and Horwitz, L. P.: A Procedure for the Diagonalization of Normal Matrices. J. Assoc. Comp. Mach., vol. 6, 1959, pp. 176-195.

16. Goldstine, H. H., Murray, F. J. and Von Neumann, John: The Jacobi Method for Real Symmetric Matrices. J. Assoc. Comp. Mach., vol. 6, 1959, pp. 59-96.

17. Goldstine, H. H., and Von Neumann, John: Numerical Inverting of Matrices of High Order. II. Proc. Amer. Math. Soc., vol. 2, 1951, pp. 188-202.

18. Gregory, R. C.: Computing Eigenvalues and Eigenvectors of a Symmetric Matrix on the ILLIAC. Math. Tab. Washington, vol. 7, 1953, pp. 215-220.

19. Henrici, Peter: The Quotient Difference Algorithm. Nat. Bur. Standards Appl. Math. Ser., vol. 49, 1958a, pp. 23-46.

20. Henrici, Peter: On the Speed of Convergence of Cyclic and Quasicyclic Jacobi Methods for Computing Eigenvalues of Hermitian Matrices. J. Soc. Ind. Appl. Math, vol. 6, 1958b, pp. 144-162.

21. Hestenes, M. R.: Determination of Eigenvalues and Eigenvectors of Matrices. Nat. Bur. Standards Appl. Math. Sec. 29, 1953, pp. 89-94.

22. Hestenes, M. R., and Karush, William: A Method of Gradients for the Calculation of the Characteristic Roots and Vectors of A Real Symmetric Matrix. J. Res. Nat. Bur. Standards, vol. 47, 1951, pp. 45-61.

23. Householder, A. S.: Generated Error in Rotational Tridiagonalization. J. Ass. Comp. Mech., vol. 5, 1958, pp. 335-338.

24. Householder, A. S.: On Deflating Matrices. J. Soc. Ind. Appl. Math., vol. 9, 1961, pp. 89-93.

25. Householder, A. S.: The Theory of Matrices in Numerical Analysis. Blaisdell, New York 1964.

26. Householder, A. S., and Bauer, F. L.: On Certain Methods of Expanding the Characteristic Polynomial. Numer. Math., vol. 1, 1959, pp. 29-37.

27. Jacobi, C. G. J.: Üher Ein Leichtes Verfahren die in der Theorie der Säcularstörungen Vorkommenden Gleichungen Numerisch Aufzulösen. Crelle's J., vol. 30, 1846, pp. 51-94.

28. Kaiser, H. F.: A Method for Determining Eigenvalues. J. Soc. Industrial Appl. Math., vol. 12, 1964, pp. 238-248.

29. Kincaid, W. M.: Numerical Methods for Finding Characteristic Roots and Vectors of Matrices. Quart. Appl. Math., vol. 5, 1947, pp. 320-345.

30. La Budde, C. D.: Two New Classes of Algorithms for Finding the Eigenvalues and Eigenvectors of Real Symmetric Matrices. J. Assoc. Comp. Mach., vol. 11, 1964, pp. 53-58.

31. Lanczos, C.: An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. J. Res. Nat. Bur. Standards, vol. 45, 1950, pp. 255-282.

32. National Bureau of Standards: Simultaneous Linear Equations and the Determination of Eigenvalues. Nat. Bur. Standards Appl. Math., Ser. 29, 1953.

33. National Bureau of Standards: Further Contributions to the Solution of Simultaneous Linear Equations and the Determination of Eigenvalues. Nat. Bur. Standards Appl. Math. Ser. 49, 1958.

34. Von Neumann, J., and Goldstine, H. H.: Numerical Inverting of Matrices of High Order. Bull. Amer. Math Soc., vol. 53, 1947, pp. 1021-1099.

35. Ortega, J. M.: On Sturm Sequences of Tridiagonal Matrices. J. Assoc. Comp. Mach., vol. 7, 1960, pp. 260-263.

36. Ortega, J. M., and Kaiser, H. F.: The L-L^T and QR Methods for Symmetric Tridiagonal Matrices. Computer J., vol. 6, 1963, pp. 99-101.

37. Parlett, B.: The Development and Use of Methods of LR Type. AEC Research and Development Report No. NYO - 10, Courant Institute of Mathematical Sciences, New York City, 1963, pp. 429.

38. Rutishauser, H.: Bestimmung der Eigenwerte und Eigenvektoren Einer Matrix mit Hilfe des Quotienten-Differenzen-Algorithmus. ZAMP, vol. 6, 1955, pp. 387-401.

39. Rutishauser, H.: Der Quotiententen-Differenzen-Algorithmus. Birkhäuser, Basel/Stuttgart, 1957.

40. Rutishauser, H.: Solutions of Eigenvalue Problems with the LR-Transformation. Appl. Math. Ser. Nat. Bur. Standards, vol. 49, 1958, pp. 47-81.

41. Rutishauser, H.: Deflation bei Bandmatrizen, ZAMP, vol. 10, 1959, pp. 314-319.

42. Rutishauser, H.: Über Eine Kubisch Konvergente Variante der LR-Transformation. ZAMP, vol. 40, 1960, pp. 49-54.

43. Rutishauser, H.: On Jacobi Rotation Patterns. Proc. A.M.S. Symposium in Appl. Math., vol. 15, 1963, pp. 219-239.

44. Rutishauser, H., and Schwarz, H. R.: Handbook Series Linear Algebra. The LR Transformation Method for Symmetric Matrices. Numer. Math., vol. 5, 1963, pp. 273-289.

45. Stiefel, E. L.: Kernel Polynomials in Linear Algebra and their Numerical Applications. Nat. Bur. Standards, Appl. Math Ser., vol. 49, 1958, pp. 1-22.

46. Strachey, C., and Francis, J. G. F.: The Reduction of A Matrix to Codiagonal Form by Eliminations. Computer J., vol. 4, 1961, pp. 168-176.

47. Tausky, O.: Contributions to the Solution of Systems of Linear Equations and Determination of Eigenvalues. Nat. Bur. Standards, Appl. Math. Ser., vol. 29, 1954.

48. Wayland, Harold: Expansion of Determinantal Equations into Polynomial Form. Quart. Appl. Math., vol. 2, 1945, pp. 277-306.

49. Wilkinson, J. H.: The Calculation of the Eigenvectors of Codiagonal Matrices. Computer J., vol. 1, 1958a, pp. 90-96.

50. Wilkinson, J. H.: The Calculation of Eigenvectors by the Method of Lanczos. Computer J., vol. 1, 1958b, pp. 148-152.

51. Wilkinson, J. H.: Householder's Method for the Solution of the Algebraic Eigenvalue Problem. Computer J., vol. 3, 1960, pp. 23-27.

52. Wilkinson, J. H.: Note on the Quadratic Convergence of the Cyclic Jacobi Process. Numer. Math., vol. 4, 1962, pp. 296-300.

NASA-Langley, 1967 —— 24 CR-983